SRI International

A WEAK LOGIC OF KNOWLEDGE AND BELIEF: Epistemic and Doxastic Logic for the Yuppie Generation

Technical Note 359

By: David Israel
Senior Computer Scientist

Artificial Intelligence Center
Computer Science and Technology Division

This research was supported in part by the United States Air Force Office of Scientific Research under contract No. F49620-82-K-0031 and in part by a gift from the System Development Foundation.

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
SEP 1985		**00-09-1985 to 00-09-1985**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
A Weak Logic of Knowledge and Belief: Epistemic and Doxastic Logic fo the Yuppie Generation	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **SRI International,333 Ravenswood Avenue,Menlo Park,CA,94025**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**32**	

Table of Contents

A WEAK LOGIC OF KNOWLEDGE AND BELIEF:

Epistemic and Doxastic Logic for the Yuppie Generation[1]

David Israel
Artificial Intelligence Center
SRI International[2]

1. INTRODUCTION

Modern modal logic begins with the work of C. I. Lewis early on in the present century [Lewis 18]. We can think of Lewis thinking to himself as follows: "Well, I can't analyze the notions of metaphysical or logical possibility and necessity; but I can sure formulate alternative axiomatizations of such notions. I can then compare and contrast such axiomatic systems and see what I learn." Thus were born the Lewis Systems, S1-S5, axiomatizing increasingly strong conceptions of necessity.[3]

Another 40 or so years went by before the purely axiomatic approach was properly systematized and rendered fit for human consumption. In current lore, a certain axiomatic system, K, is central.[4] The standard presentation of K consists of infinitely many axioms plus one axiom scheme and two rules of inference. In particular, with 'L' being read as "necessarily" or "it is necessary that"; 'M', as "possibly" or "it is possible that", K is as follows:

I: all classical tautologies

II: L(p --> q) --> (Lp --> Lq)

---

[1] I will now show off almost all the Greek I know: "epistemic" has to do with knowledge; "doxastic", with belief. So in what follows we shall have to do with logics of knowledge and belief.

[2] This research was supported in part by the United States Air Force Office of Scientific Research under Contract No. F49620-82-K-0031 and in part by a gift from the System Development Foundation.

[3] The little story just told is a fable. Lewis was really interested in different conceptions of implication or the conditional--not in varying conceptions of necessity and possibility. Of course, on one view, implication simply *is* validity or *necessity* of the material conditional; so we can translate Lewis's writings on the varieties of implication into writings on varieties of necessity. This translation scheme is now almost universally applied. Note, if one does not apply this scheme, and instead reads Lewis neat, the proper line of descent from Lewis goes mainly through Ackermann's work on 'strenge Implikation' to the work of Anderson-Belnap on entailment. See [Anderson and Belnap 75].

[4] The "K" is for Kripke, although credit for focussing on a notion of normality under which K is the minimal normal modal logic must be shared with E.J. Lemmon [Lemmon 77]. See below on normality.

R1: If l-p and l-(p --> q), then l-q *modus ponens*

R2: If l-p, then l-Lp *necessitation*

The standard practice is to take K as the base theory and consider extensions. Four such extensions have figured prominently in the literature.

T: K + **Lp --> p**

$S4$: T + **Lp --> LLp**

B: T + **MLp --> p**

$S5$: T + **MLp --> Lp**

In all of these logics, possibility and necessity are duals; that is, in all of them "Lp" is provably equivalent to "-M-p" and "Mp" to "-L-p". Thus they can all be with only one primitive modal operator ('M' or 'L')--its dual ('L' or 'M', respectively) being introduced by definitional abbreviation.

Just to confuse the reader, I shall spend a little time on alternative systems of nomenclature for modal systems. First, and least annoying, T is also referred to as M. Now then, look at the characterization of, say, M. (Just testing.) M is presented as K plus one axiom schema. That schema is also often referred to as **T**--though never, I think, as **M**. Thus T, the system, just is K + **T**, the schema. This particular annoyance, or variants of it, recurs. The schema, which when added to K + **T** yields $S4$, is called **4**; that, which when added to K + **T** yields B, is **B**. Finally, the $S5$ schema is **E**. The scorecard looks likes this:

T = K + **T**

$S4$ = K + **T** + **4**

B = K + **T** + **B**

$S5$ = K + **T** + **E**

In the remainder of this paper, I shall adhere to the conventions manifested on the right hand side of these equations; thus, I shall be looking at systems that are presented as K + **X**, **X** the unknown.

2. ON AXIOMATIZING KNOWLEDGE AND BELIEF

To return to the main line: these four standard modal logics were meant to formalize different conceptions of necessity and possibility. They were not meant to cast any light on the notions of knowledge or belief--or on different conceptions of knowledge or belief. Indeed, what *a priori* reason is there to believe that any of these standard logics of **necessity** are appropriate logics of **knowledge** or **belief**? Whatever the answer to that, Hintikka [Hintikka 62] gave people lots of reasons *a posteriori* to think that (1) $K + \mathbf{4}$ was an appropriate logic for knowledge and (2) $K + \mathbf{E}$ was an approriate logic for belief. (Note: $K + \mathbf{E} = S5 - \mathbf{T}$. This is sometimes called "weak S5.")[5]

The response to Hintikka's work was quite stunning--as these things go; and as they went, no one paid much mind to the logic of belief. The focus was squarely on knowledge--to philosophers, at any rate, the more interesting and more discussed notion. Many attempts at conceptual analysis of the notion of knowledge had been made; none had met with exactly universal acceptance. So why not go Lewis's route: don't analyze, axiomatize? **Especially now!!**

Why especially now?? Because in the interim (1918 to 1962), logicians had come up with model-theoretic tools for a variety of modal logics--including our four standard ones. (It was a number of years before it was clear how wide a variety this was.) Further on, we shall look at the main ingredients of the now standard model theoretic treatment; for now, it suffices to note its very existence and to note that its existence played a large part in the excitement surrounding Hintikka's work.[6]

Still, there was trouble in the new paradise. It came in two quite independent forms. First, there was the problem of *logical omniscience*, so-called. Then, there were problems about introspection. As for the first problem; it is easy to prove that K by itself--with 'K' substituted for 'L', of course--guarantees both that every classical tautology is known and that knowledge is closed under classical tautological consequence. The latter means that if S' follows tautologoulsy from S and if it is known that S, then it is

---

[5]The sharp-eyed reader might have guessed that there were more notational headaches ahead. However it came to be that 'L' got associated with "it is necessary that" and 'M' with "it is possible that", it was only to be expected that 'K' would be used for "it is known that" and 'B' for "it is believed that". But now 'K' stands for both an axiomatic system and a modal operator; 'B', for a modal system, an axiom scbema, *and* a modal operator. Context, together with my convention of *italicizing* system names and **boldfacing** schema names, will disambiguate. By tbe way, I trust that it is clear that knowledge ('K') and belief ('B') are not duals. From "it is not believed that it is not the case that p", "it is known that p" does not follow; nor vice versa. Nor should one infer from "it is not the case that it is known that it is not the case that p" to "it is believed that p"; or vice versa.

[6]More fabulating; Hintikka's original work was not done within the then new model theoretic framework; the "semantic" machinery was, rather, syntactic and proof-theoretic. In later versions, Hintikka did adopt the new standard.

known that S'.[7] Idealization is fine, indeed necessary in any science; but surely this is going too far with a fine thing.

The second set of problems had to with what one should add to K + **T** for knowledge or to plain old K for belief. (Remember that, sad to say, we can't allow ourselves **T** for belief.) Hintikka spends a good deal of time arguing for the inclusion of **4**, at least for knowledge. Many thought that this was too strong a requirement. He also argued against the inclusion, again for knowledge, of **B** and **E**. Here the consensus was with him. Questions were raised about belief as well. Could one believe that p without believing that one believed that p? That is, should one add **4** to K? Could one not believe that p without believing that one did not believe that p? That is, should one add **E** to K?[8]

[7] K guarantees more: if S is any theorem of K--it need not be a classical tautology– then it is known that S; this is just what the rule of necessitation yields. Mutatis mutandis for closure under consequence; think of it as closure under K-consequence.

[8] A word in explanation of the grotesqueries of logician's Englisb. "Scott doesn't believe that p" is ambiguous. It can be understood to mean that Scott–for whom, see below–believes that not-p or to mean simply that it is not the case that he believes that p. Scott might not have any fixed opinion as to whether p. In what follows, it is crucial that these two readings be distinguished; the ugly way, deploying negation only as a sentence-level operator in the guise "it is not the case that", is the way for me. To make matters worse, I refuse to countenance any natural dual for either "knows" or "believes", either 'K' or 'M'. It is nice that "necessarily" and "possibly" are (arguably) lexicalized duals; thus, we don't have to keep writing down things like "it is not the case that it is necessary that it is not the case that..". We can write instead "it is possible that.." But not only aren't "knows" and "believes" duals, neitber has a natural, lexical dual. So there will be lots of ugly tbings like "it is not tbe case that Scott believes that it is not the case that Scott believes that Scott believes that p." Sorry.

3. INTRODUCING SCOTT AND KIMBERLY

To fix ideas, let's imagine a subject. To fix our perhaps sexist imaginations, let's imagine two subjects, Scott and Kimberly. So, in what follows 'K' is to be read as "Scott (Kimberly) knows that..." and 'B', as "Kimberly (Scott) believes that..." The formalisms I will be discussing are all of the single subject variety. I shall have nothing to say about the multisubject versions being studied by researchers in theoretical computer science interested in distributed systems [Halpern and Moses 84].[9]

Scott and Kimberly are, of course, terrifically bright; but are they logically omniseient? Why not make their mommies and daddies happy by assuming that they are. This deeision also makes me happy; for a mixture of tactical and technical reasons, I think it useful to retain K as our base theory. For alternatives to this, see [Fagin and Halpern 85].

In any case, unrestricted necessitation is out for any applied epistemic or doxastic logic. Imagine that we are interested in some set of putative facts and in what Kimberly knows/believes about them. One such fact might be that South San Francisco calls itself "The Industrial City." We add a sentence expressing that fact as an axiom in an applied modal logic; but, we don't want to apply necessitation. We don't want to infer, that is, that Kimberly knows/believes that south San Francisco calls itself "The Industrial City." What does a classy kid like Kimberly care about a place like South San Francisco? We shall have to simply add particular axioms about what Kimberly does (or does not) know/believe about the situation in question; or, better, those facts are part of the situation in question.

The worries about introspection are horses of another color. It is those that I am going to try to honor. One crucial consideration here is sociological. Yuppies simply are not very introspective; they're much too busy networking and consuming to be self-reflecting. The pale cast of introsection surely gets in the way of having good, trendy, expensive fun; one can't get all there is out of driving one's BMW if one is paying attention to one's own thought processes—as opposed to the impression one is making on others of one's kind, etc., etc. Another consideration is a fondness on my part for weak noncommittal systems to which one can add strength--and bold committments—as one one wishes.

---

[9]Single subject epistemic/doxastic logics will have two unary modal operators, 'K', 'B', each with a subscript suppressed but both fixed and understood. That is, one is to fix a subject, say Scott, and read 'K' as "Scott knows that.." Of course, if one assumes--as I shall--that all Yuppies are in the relevant respects indistinguishable, one can imagine oneself working with a schematic modal operator, an operator whose subscript is a schematic letter whose substitution instances are singular terms for Yuppies; e.g. names like "Scott," "Kimberly"—not e.g., "Harvey," "Alice."

4. ON KNOWLEDGE.

As noted above, Hintikka argued strenuously for the epistemic version of **4**: the thesis that if one knows, one knows that one knows. People attacked this position; Hintikka relented, as well he should have. Most of the bad arguments for skepticism--that is, most of the arguments--have turned on tricking the ingenuous into accepting the thesis that if one knows, one knows that one knows and then arguing that one doesn't know that one knows. Let us suppose that knowledge requires either justification on the knower's part or a "proper" etiology for the belief, e.g. a suitable placement on the knower's part with respect to the fact known (e.g., standing in the right kind of causal relation to it).[10] Surely either of these requirements can be met without the knower's knowing that they're met. Indeed, surely we might sometimes be argued into accepting unreasonably high standards on knowing--so high that though we know, we not only don't know that we know, we actually believe (falsely) that we don't know. Of course, if we're sufficiently gullible, such arguments might even get in the way of the controverted belief (our knowledge of which was in question), so that we cease to know that p because we have (foolishly) ceased to believe it.

For Hintikka's original epistemic logic we can prove that the addition of the axiom schema **4** is equipollent with the addition of the following rule of inference:

RKK: If ⊢(Kp --> q), then ⊢(Kp --> Kq)

For one direction of the proof of equipollence; we have ⊢(Kp –> p) (by **T**), whence by **RKK**, we have ⊢(Kp –> Kp), whence, by **RKK** yet again, ⊢(Kp –> KKp). (The other direction is left as an exercise for the reader.) Imagine that whether Scott knows that p is up for grabs, and let q be any old sentence the truth of which is sufficient for the falsity of the claim that Scott does know that p. Now reason contrapositively and apply **RKK**. To wit;

(q --> -Kp); so (Kp --> -q); so--by **RKK**--(Kp --> K-q)

This may seem innocuous; but it isn't. In order to know that p, poor Scott must know the falsity of anything whose truth rules out his knowing that p. This is precisely the sceptic's trick. Get someone to accept this requirement, and it won't be hard to get that same someone to doubt that anyone knows anything. For the requirement certainly seems to amount to this: if Scott does know that p, then he

[10] This supposition encompasses the supposition that knowledge is not just true belief. Much of the recent AI and computer science literature *seems* to suppose that knowledge is just true belief. But it isn't.

knows the falsity of anything whose truth would rule out his knowing that p. We might say, then, that Scott, in knowing that p, must be in a position to disregard all further evidence with respect to–i.e., in a position to rule out any and all counterpossibilities. But Scott is almost never in a position to disregard all further evidence; so Scott almost never knows anything.

Now all this may be an abuse of the thesis that if one knows, one knows that one knows. (Though I should note that the argument just given is used by Hintikka himself in his–somewhat reluctant–recantation of the axiom. See [Hintikka 70.] Still, I see no reason to accept the thesis. Indeed, I see no reason to accept even the claim that if one knows one believes that one knows. If one does believe that one knows that p, one might be said to be certain that p. At least, that is how the philosopher G. E. Moore characterized certainty. Provisionally accepting this characterization, I want to say that one can know that p without being certain that p.

Hintikka also spent time arguing against the epistemic version of **B**:

(-K-Kp --> p)

This says that if it is not the case that Kimberly knows that it is not the case that Kimberly knows that p, then p. This is truly bizarre; a little "introspective ignorance" on Kimberly's part about the scope and limits of her knowledge is going an awful long way. (I suppose her parents–dabbling in epistemic logic–*might* look favorably on this schema; but surely cooler heads would ultimately prevail.) Ruling out **B**, while accepting $K + \mathbf{T}$, as Hintikka does, provably rules out accepting the epistemic version of **E**:

(-K-Kp --> Kp)

That's no great price to pay since the epistemic version of **E** seems wildly too strong. (Thus, by simple transformations, this yields that if one does not know that p, then one knows that one does not know that p. Would that life were so neat!)

One last word about knowledge and the so-called introspective axioms. I noted in passing that knowledge certainly seems to be more than just true belief. In particular, it seems to require that the belief be justified or that it (and the believer?) stand in some special–perhaps causal–relation to the fact. Eternally controversial issues in the philosophy of knowledge lurk. Let them lurk; it suffices for my purposes to point out that if one buys some version of the second, "causal," account of knowledge–as I am inclined to do–then the knowledge that one knows need not be, in any clear sense, introspective–beyond the bare minimum of knowing that one believes that p, if one does. Rather what one must know

to know that one knows that p is that one (or one's mental state of believing that p) stands in the right kind of causal relation to the fact that p. This might involve knowledge about one's sensory apparatus, as well as knowledge about more fully external features of the situation. But this is surely not introspective knowledge at all. (Indeed, there are, I think, similarly external or objective readings of some versions, at least, of the justification story--readings which turn justification-based accounts into "causal" accounts.)

In sum: with respect to the axioms governing the "K" operator, I opt for minimality (modulo some version--restricted or not--of "logical omniscience"). That is, I opt for the epistemic version of K + **T**. The modal core of our epistemic logic is just the modal core of K:

II': **K(p --> q) --> (Kp --> Kq)**

R2': **If l-p, then l-Kp**

5. ON BELIEF

As to belief: if no one else and if no one earlier, Freud should have taught us that we don't always know our own minds. Indeed, we can't always believe our minds are as they, sad to say, are. We can believe without believing that we believe; so much for the doxastic version of **4**. We can also not believe that we do not believe that p and still not believe that p. That is to say, the doxastic version of E seems false:

(-B-Bp --> Bp)

Likewise the doxastic version of **B**, which like its epistemic counterpart seems crazed—only more so:

(-B-Bp --> p)

If Kimberly doesn't believe that she doesn't believe that p, then p. This is megalomania, even in someone as spoiled as Kimberly is likely to be.

A last word on the standard "introspective" axioms for belief: it can seem as though one's beliefs about one's own beliefs will typically be vouchsafed one by introspection. This seeming gets weaker when one considers past--or future—beliefs of one's own. Certainly for the past, there's memory; but memory of what? Of one's past mental states or of one's past actions? Thus, we often reason as follows: I must have believed that p; for consider what I did. Independent of Freud, et al., I think there are good reasons for doubting the extent of one's introspective access to one's own current beliefs. Some of these reasons have to do with the nature of the objects of belief; some, with the nature of believing as a state.[11] I'm not going to rehearse these here. Instead, I will simply present another scorecard:

AXIOMS RELATING BELIEF AND KNOWLEDGE THAT I ACCEPT

Kp --> Bp

AXIOMS RELATING BELIEF AND KNOWLEDGE THAT I DO NOT ACCEPT

Bp --> BBp

---

[11] I will return to the question of the objects of belief, albeit briefly, below.

Bp --> KBp

Kp --> BKp

Kp --> KKp

6. ON LIMITING INTROSPECTION FOR BELIEF

So, what axioms *do* I want for belief, at least for the beliefs of such as Scott and Kimberly. First, let me remind the reader that, however taken we may be with these Young Upwardly Mobile Professionals, they are not infallible. We cannot allow them the doxastic version of **T**: Bp --> p. We might, though, grant them a consistency condition--this comes in especially handy for those whose beliefs are closed under classical tautological consequence. The condition in question is that if Kimberly believes that p, then she does not believe that it is not the case that p. This is a doxastic version of a schema called **D**.

(D): **(Bp --> -B-p)**

The 'D' is for "deontological" or "deontology." (More Greek.) Deontology is the study of the logic of obligation. The crucial operator there is "it is obligatory that...". It does have a dual: "it is permissible that...". Note that just as we cannot, alas, have a doxastic version of **T** for reasons of fallible belief; so too can we not have a deontological version--for reasons of fallible mores. But we do have it that if it is obligatory that p, then it is permissible that p. That is, if it is obligatory that p, then it is not obligatory that it not be the case that p. This last is just **D**. So **D** is oft regarded as the characteristic deontological axiom. It is, of course, obvious that **D** is a theorem of K + **T**. Is **T** a theorem of K + **D**? We must hope not, for then by granting consistency, we will let in the unacceptable infallibility. How can one tell?

There is one sure way to tell that something *is* a theorem of a given system--prove it within the system. In general, only infinite patience will avail if one wants, obversely, to show of a sentence that is not a theorem of some system that it is not. Even for decidable systems--and all the logics I will be discussing here are decidable--"direct" proofs of nontheoremhood are really out.

6.1. *Model Theory of Intensional Logics*

Model theory to the rescue! The model theory of modal logics is good for at least two things: (1) proving in the semantic metatheory that such-and-such is a theorem of so-and-so and (2) proving in the metatheory that such-and-such other thing is not. Logicians, generally, aren't sufficiently silly as to want to work *within* a given formal system; they prefer to work on the outside, using whatever tools are appropriate, to prove things *about* the formal system. This is what Kripke et al. allowed logicians to do with respect to modal logics. The key to Kripke's analysis lies in the introduction of modal models; triples <S, R, v> where S is any nonempty set, R is a relation on S, i.e. a subset of S X S, and v is a value assignment meeting standard conditions for standard sentences and the following condition for modal

sentences. Using 'L' now as the strong, necessity-style operator and, the redundant but useful, 'M' as its dual:

L: For any wff. p, and any s in S, v(Lp, s) = T
if v(p, s') = T for every s' in S s.t. sRs';

otherwise v(Lp, s) = F.

M: For any wff. p and any s in S, v(Mp, s) = T
if there is at least one s' in S such that sRs' and such that v(p, s') = T;

otherwise v(Mp, s) = F.

So the necessity operator is akin to the universal quantifier; its dual, the possibility operator, akin to the existential quantifier. R enters the above as a parameter--as does S, for that matter. What Kripke, et al. showed was that one could ring changes in the nature of R and thereby yield modal models appropriate to different modal logics. One way to think about this is to ignore the value assignments and think of duples: <S, R>, S and R as before. Call such things *frames*, and go on like this: a formula is valid on a frame just in case it is valid in every model based on that frame--letting v vary. Finally, say that a modal system is **characterized** by a class of frames if all and only its theorems are valid on every frame in that class. *Voila*, different modal systems are characterized by different classes of frames, the difference residing precisely in the conditions on R.[12]

Now, as to why *K* is called the **minimal normal** modal logic. Simple, *K* imposes no restrictions on R at all--not even that it be nonempty. So much for minimality. As for normality; here, it's what's **not** in S, as opposed to the nature of R, that counts. Call a subset Q of S **nonnormal** if for every q in Q, every wff. p, and every v, v(Mp, q) = T and v(Lp, q) = F. If Q is empty, then S is **normal**. At

[12]Before getting down to some cases, I should bring to the reader's attention my use of the letter 'S', in place of 'W'. The foregoing story is often glossed as follows: let S be a set of possible worlds, and R a relation of relative accessibility between worlds. Necessity is truth in all accessible possible worlds; possibility, truth in at at least one. This gloss is just that: the heuristic of thinking of the members of S as possible worlds is, of course, no part of the formal development. Worse, it can be seriously misleading. Don't, dear reader, let it mislead you. In (almost) the immortal worlds of Brendan Behan:

Don't muck about

Don't muck about

Don't muck about

with

Possible worlds

For an alternative semantic picture, see [Fagin and Vardi 85].

nonnormal "indices" or "points of evaluation" (each much more appropriately neutral than "possible world"), anything is possible and nothing is necessary. Sounds like fun.[13]

Restricting ourselves to extensions of K, we can speak either of the characteristic condition on R associated with a given schema **X** or of that associated with the system that consists of K + **X**. I shall speak in the former mode. So here's another scorecard:

SCHEMA	CONDITION ON R
T	(s)(s R s) *reflexivity*
4	(s, t, u)(s R t & t R u --> s R u) *transitivity*
B	(s,t)(s R t --> t R s) *symmetry*
E	(s, t, u)(s R t & s R u ---> t R u) *Euclidean condition*
D	(s)(Et) (s R t) *seriality*

It is now obvious that if R is reflexive it is serial and just as obvious that R can be serial without being reflexive. So, K + **T** yields **D**; but K + **D** does not yield **T**. We're safe; Scott and Kimberly can be logically omniscient and consistent, at least with respect to their beliefs, without being infallible.

8.2. *Some Applications.*

Now that we have some tools at our disposal, there are other conditions we might want to consider. Scott and Kimberly, after all, think mighty highly of themselves; perhaps, although they are not infallible, they think they are. One expression of this unseemly immodesty—nay, arrogance—is **U**:

U: **(B(Bp --> p))**

We shall reject **U**. To give it its model theoretic due; there corresponds the following nameless characteristic condition on R:

---

[13]Nonnormal worlds–frames containing such–enter into the semantics of Lewis's *S1* - *S3*–the differences among these being correlated with differences in the accessibility relationship. No one has ever taken these systems very serioulsy as logics of necessity and possibility. Of course, if I may remind the reader of the fabulous nature of the fable with which I began, they were not meant to be such. I should also note that frames with "impossible" possible worlds have been looked to for a way of handling, within modal logic, the problems of logical omniscience. I will have nothing to say about such attempts in *this* essay. See [Hintikka 75].

$$(s,t)(s\ R\ t \dashrightarrow t\ R\ t)$$

One can show that **D** and **U** are independent. To show that **D** does not yield **U**, consider the following frame:

$$\mathbf{S} = \{s,\ t\} \qquad \mathbf{R} = \{\langle s,\ t\rangle,\ \langle t,\ s\rangle\}$$

Here R is *serial*, but not **U**-ish. (Does it *look* **U**-ish?) To get a frame which satisfies **U** but not **D** is but a moment's work:

$$\mathbf{S} = \{s,\ t\} \qquad \mathbf{R} = \{\langle s,s\rangle\}$$

Note that this frame is not *reflexive*, that is, not **T**-ish. So, **D** and **U** are independent and both are weaker than **T**. Indeed, K + **D** + **U** is a system strictly weaker than K + **T**, for note the following:

$$\mathbf{S} = \{s,\ t\} \qquad \mathbf{R} = \{\langle s,\ t\rangle,\ \langle t,\ t\rangle\}$$

6.3. *A Few Doxastic Paradoxes*

Before leaving **U** behind, I should note the connection between it and the so-called "Paradox of the Preface." **U**, as noted, is too much; not even Scott and Kimberly believe they are infallible. So, both Scott and Kimberly believe that one of their beliefs is false. But then not all of their beliefs could be true.

Take Scott. He's a reasonable fellow, as Yuppies go. He believes that at least one of his beliefs is false. That is, not only does he not conform to the self-regarding standards of **U**; he positively repudiates same. Now either this belief in his own fallibility—call it **non-U**--is false or not. If it is false, then at least one of his beliefs is false--viz. **non-U**; so **non-U** is true. And, of course, if **non-U** is true, then at least one of his (other) beliefs is false. So whether **non-U** is true or false; it is true. So Scott's belief that at least one his beliefs is false must be true; so at least one of his beliefs is false. **non-U** is fated to be true.

Let's go more slowly here. Let's assume that the "range" of **non-U** does not include **non-U** itself. That is, Scott believes that at least one of his beliefs other than **non-U** is false. Suppose, for simplicity's sake, that Scott has finitely many other such beliefs: p, q, r,...Suppose, further, that Scotts's beliefs are closed under (finite) adjunction. (This, by my lights, is likely to be a wild supposition; in general, the supposition of unrestricted adjunction--for any attitude--is an extremely dubious one. This is one reason for being dubious about K.) Scott, then, believes the conjunction of p with q with r...But he also believes **non-U**; this is to believe: either not-p or not-q or not-r or... But these two beliefs are inconsistent. Neither

one of them is **non-U**; at least one must be false; so **non-U** is true. Of course, although the second of the two beliefs--the disjunction of the negations of the conjuncts of the first--is not itself **non-U**, it records--in the context of the finitely many other beliefs conjoined in the first--the effect of believing **non-U**.[14]

The situation is even more baroque if we put **non-U** back into the pot of Scott's beliefs. Indeed, it is the situation as outlined two paragraphs ago. The supposition that **non-U** is not true leads immediately to the conclusion that it is true. But we needn't stop there. Return to the troublesome case where all of Scott's other beliefs are true. If **non-U** cannot but be true, then it is true. But then all Scott's beliefs are true, after all. But then **non-U** is false, after all. Something is wrong somewhere.

What seems to be wrong is that Scott, no matter how hard he tries, can't successfully believe--either truly or falsely--that at least one of his beliefs is false *unless* one of his other beliefs is false. In which case, of course, no matter how hard he tries, Scott can't help but believe truly that at least one of his beliefs is false--if he believes it at all. Again if Scott were ever successfully to believe the *first-person* version of the negation of **non-U**, then *that* belief would be guaranteed to be true. The first-person version of **U** is a bit much--even for Scott; the third person version, a bit much even for his parents. The third person-version of **non-U** seems just fine--surely it is not the case that Scott believes that if Scott believes that p, then p. Finally, the first-person version just cannot be falsely believed.[15]

So much for **U**. There is another paradox about: to wit, Moore's paradox. (Arguably the first pragmatic paradox to be remarked upon.) Let's pick on Kimberly this time. Kimberly, poor lass, has false beliefs. So we will have occasion to say such things as:

Kimberly believes that p; but it is not the case the p.

Moreover, Kimberly is not omnidoxastic; there are truths she simply does not believe. (I will speak of the trait of believing all the truths there are as *omnidoxasticity*.) So we will have occasion to say such things as:

p; but Kimberly doesn't believe that p.

Kimberly, moreover, believes that she has false beliefs--if you don't believe me, advert to the above and ask Scott. But, notice how odd it would be for her to say:

---

[14] I shudder with this talk of conjoining and disjoining beliefs; still, it's a convenient shorthand -- but *for what?*

[15] This discussion of the Paradox of the Preface is just a retelling of a tale told, in Polish notation, by A.N. Prior. [Prior 71.] The Paradox was first noticed by D.C. Makinson.

I believe that p; but it is not the case the p.

It is perhaps odder even for her to come out with the first-person denial of omnidoxasticity:

p; but I don't believe that p.

G. E. Moore first pointed out the paradoxical character of the first-person versions of what, in third-person forms, are completely innocuous things to say. I have said that the schematic letters that come with of our 'B' and 'K' operators were to have singular terms for subjects as substutition instances. "I" is such a singular term, but a very special one. Note that even if your name were Kimberly--and you were alone in being so named--it could be perfectly nonparadoxical for you to say:

p; but Kimberly doesn't believe that p.

You might, after all, not know your own name, might not--in this sense--know who you are. Without going much more deeply into problems about indexicals and quasi-indexicals, we cannot really go very deeply into Moore's Paradox; so, in what follows, I am going to be playing a little fast and loose. I am going to assume that Scott knows who he is, at least in so far as he knows that he is (the one and only) Scott--the one and only person named 'Scott', mutatis mutandis for Kimberly. In this respect, I follow Hintikka's lead.

6.4. *Moore's Paradox and the Schema* Y

To return to the main line, the key here is the following schema:

(p & -Bp)

We cannot rule it out by ruling in its denial:

-(p & -Bp)

for that is equivalent to the wholly unacceptable **O**:

(p --> Bp) *omnidoxasticity*

Note that the negation of our version of *"I believe that p; but it is not the case that p"*. is the equally unacceptable infallibity axiom **T: (Bp --> p)**.

What we want to rule in is

(-B(p & -Bp))

It is not the case that Scott believes both that p and that it's not the case that he (Scott) believes that p. This is equivalent to:

(-B-(p --> Bp))

It is not the case that Scott believes that it is not the case that if p, then Scott believes that p. That is, though Scott does not believe in his own omnidoxasticity, it is not the case that he believes that it is not the case that he is omnidoxastic. Another perspective on this dark saying is vouchsafed us by distributing "B" over "&" in the earlier version:

(O'): (-(Bp & B-Bp))

It is not both the case that Scott believes that p and that he believes that it is not the case that he believes that p.

This last raises the question of **U** again. I have simply assumed that Scott does not believe he is infallible. So we do not accept

(U): (B(Bp --> p))

But we can deny that Scott believes the negation of the infallibility schema. We can allow

(-B-(Bp --> p))

This is equivalent to

(-B(Bp & -p))

It is not the case that Scott believes both that he believes p and that it is not the case that p. Or distributing "B" over "&":

(U'): (-(BBp & -Bp))

It is not the case both that Scott believes that he believes p and it is not the case that he believes p.

Perhaps the basic drift is now clear. I do not want to buy the standard "axioms of introspection"; **not even for belief**. Rather, the logic of belief I am proposing is generated by the intuition that what one wants is that one's subjects--Scott, Kimberly--not be stuck with certain kinds of false introspective beliefs. So, I propose that they not make certain kinds of mistaken self-acsriptions of belief; thus, that they not both believe that p and believe that they do not believe that p. Again, they should not both not believe that p and believe that they believe that p. To grant this freedom from error is already a generous gesture of idealization on my part; but, of course anyone as blithely unconcerned with "logical omniscience" as I cannot blanch at idealizing. Still, it is a much weaker form of idealization than guaranteeing oodles of true self-ascriptive beliefs. Yuppies, remember, don't introspect much.[16]

The key idea in the above might be put as follows: take a controversial schema and deny that Scott or Kimberly believes its negation. This is exactly how we got **O'** from the omnidoxastic schema **O**; and **U'** from the obnoxiously self-satisfied **U**. Let's apply this algorithm to the doxastic versons of both **4** and its converse, the unnamed

(BBp --> Bp)

Let's name this **C4**, for the converse of 4. This is not to be confused with **U**. (Though it is entailed by, it does not entail, **U**.) What we get, after the standard transmogrifications, are **4'** and **C4'**.

(4'): (-B(Bp & -BBp))

---

[16]Of course, some true self-ascriptions creep in with the logic, with *K* itself. For instance, by **R2**: ⊢(p --> p); so ⊢(B(p --> p); so ⊢B(B(p --> p)). Voila, introspection! But nothing to write home about.

(C4'): **(-B(BBp & -Bp))**

And, if one thought it more perspicuous:

(4'): **(-(BBp & B-BBp))**

(C4'): **(-(BBBp & B-Bp))**

Yuppies may not be introspective; but they are **confident**–even about the rare introspective beliefs they may entertain. Although it is not the case that if one believes that p, then one believes that one believes that p; neither is it the case that one believes both that one believes that one believes that p and yet does not really believe that p. (That was **4'**,in case you couldn't guess.) Moreover, it isn't even true that if one believes that one believes that p, then one does believe that p. But it is true that one doesn't believe both that one believes that one believes that p and yet that one doesn't believe that p. (That's **C4'**.)

Let's look back at **O'**, the one prize we captured from our perusal of Moore's Paradox:

(O'): **(-(Bp & B-Bp))**

This is equivalent to

(-B(p & -Bp))

which is, in turn, equivalent to

(Y): **(Bp --> -B-Bp)**

If Kimberly believes that p, then it is not the case that she believes that it is not the case that she believes that p; more colloquially: if she believes that p, then she doesn't believe that she doesn't believe that p. *Nota bene*: no **real** introspection is required; rather, what is being ruled out is that Kimberly have certain kinds of false introspective beliefs.[17]

---

[17]The contrapositive of Y is (B-Bp --> -Bp). If Kimberly has any positive introspective belief to the effect that she does not believe that p, then she does not believe that p. Note the asymmetry here between negative and positive introspective beliefs. As Achilles said to the Tortoise, "That's Classical Logic".

If we add this schema, which we have dubbed **Y** for the obvious reason, we have a modal system that will yield all of **D**, **U'**, **4'**, **C4'**, and none of **O**, **U**, **4**, or **C4**.[18]

6.5. *More on* Y

Y is, of course, the doxastic version of the nameless **(Lp --> MLp)**. This latter is just an instance of **(p --> Mp)**, which is a fairly basic principle about possibility. Indeed it is just the other side of the coin from **T**. But we don't have **T** for belief; nor do we have **(p --> -B-p)**. We have **Y**.

No doubt the reader is just dying to get a gander at the characterstic condition on R associated with **Y**. Take a gander:

(Y): $(s)(Et)(s\ R\ t\ \&\ (u)(t\ R\ u \dashrightarrow s\ R\ u))$

That any frame which is **Y**-ish is *eo ipso* **D**-ish--that is, *serial*, is obvious. The converse does not hold. Consider:

S = {s,t,u} **R** = {<s,t>, <t,u>,<u,s>}

This is serial but not **Y**-ish. Thus, $s\ R\ t$ and $t\ R\ u$; but it is not the case that $s\ R\ u$. Moreover, **Y** does not yield **U**. (Remember, we don't want it to.), thus:

S = {s,t,u} **R** = {<s,t>,<t,u>,<s,u>,<u,u>}

Note that though $s\ R\ t$, it is not the case that $t\ R\ t$. Indeed this shows that **Y** does not yield the unwanted **T**.

---

[18]NOTA BENE: Craig Harrison, in a discussion of the paradox of the unanticipated examination, has argued for a modal logic of belief much like the one proposed here. See [Harrison 80]. Actually, his proffered alternative is weaker; it is essentially $K + D$. But he, too, considers the schema I have called Y. Moreover, he, too, adduces Moore's Paradox as, at the very least, a consideration . The history here is complicated. The work on what is now *Yuppielogic* began almost fifteen years ago, after Harrison's paper appeared. When I began the work, I hadn't yet read Harrison's paper. Indeed, I wasn't thinking about the paradox of the surprise exam at all. Then, as in the present essay, I ignored all issues of time and its passage; then, as in the present essay, the considerations for and against various principles had their source in Moore's Paradox, (to a lesser extent) the Paradox of the Preface, and general epistemological considerations. In fact, I think Harrison's treatment of the unanticipated or suprise exam extremely interesting, but--in the end--inadequate. He bases too much on a rejection of the theses that if one knows/believes, one knows/believes that one does. I, too, reject those, though not simply (or at all) because that decision allows one to hold that the set up in the surprise exam puzzle is a consistent one. Though Harrison treats of time indexed epistemic and doxastic operators, he doesn't do enough with them, doesn't say enough about the principles that should govern them. In any event, I hope to address the issues raised by that paradox in the future. Still, I don't want it thought that the idea of looking at intensional logics for belief and knowledge which extend K but not as far as any of the standard logics of necessity do, is either unique with, or original to, me. Harrison, and no doubt others, including Binkley [Binkley 68], got there first.

As to **(BBp --> Bp)**: Its characteristic condition is as follows:

$$\textbf{(BBp --> Bp)} \qquad (s,t)(s\ R\ t \dashrightarrow s\ R^2\ t)$$

So consider the frame:

$$\mathbf{S} = \{s,t,u\} \qquad \mathbf{R} = \{\langle s,t\rangle,\langle t,u\rangle,\langle s,u\rangle,\langle u,u\rangle\}$$

Here, $s\ R\ t$; but it is not the case that $s\ R^2\ t$. That is, there does not exist an s', s.t. $s\ R\ s'$ & $s'R\ t$.

Finally, as to **4**, **(Bp --> BBp)**:

$$\mathbf{S} = \{s,t,u\} \qquad \mathbf{R} = \{\langle s,s\rangle,\langle t,t\rangle,\langle u,u\rangle,\langle s,t\rangle,\langle t,u\rangle\}$$

$s\ R\ t$ and $t\ R\ u$; but it is not the case that $s\ R\ u$. (This particular frame is reflexive; but, of course, not all **Y**-ish frames need be.)

I assume, by the way, that it's obvious that **Y** yields neither **B** nor **E** nor **O**. The characterstic condition of this last is: $(s,t)(s\ R\ t \dashrightarrow s = t)$.

So much for the crucial negative results. Let's think positively. We've already noted that **Y**--that is, K + **Y**--yields **D**. **Y** yields **O'**, because it *is* **O'**. There are fairly straightforward direct proofs in K + **Y** of **U'**, **C4'** and **4'**.[19]

The reader may well wonder about the results of applying the above treatment to **B** and **E**. That is, what about the schemata that result by negating the believability--for such as Scott and Kimberly--of their negations? The resulting schemata are, in order, **B'**:

(-B(-B-Bp & -p))

or:

(-(B-B-Bp & B-p))

---

[19]Rather than bore the reader to tears with such proofs, I'll give hints for their construction. To prove U', simply substitute 'Bp' for 'p' in Y; to get 4' from U' is the work of but a moment, making use of the same substitution pattern as before--'Bp' for 'p'. Finally, to get C4': use axiom scheme II of K on D, put the result of that together with Y, and *Voila*, C4'.

and **E'**:

(-B(-B-Bp & -Bp))

or:

(-(B-B-Bp & B-Bp))

These are sufficiently opaque as to not be worth much worry; but, in fact, they are both theorem schemata of $K + \mathbf{Y}$.[20]

[20] Proofs left as nontrivial exercises for the reader.

7. SUMMING UP

It is time both to sum up and attempt, at least, to see *YUPPIELOGIC* from a proper perspective. Any "logic" of knowledge and belief will have to be based on idealizations. There are, at least, two orthogonal dimensions along which to idealize. One dimension is that of the the logical competence of the subject knowers/believers. The other is that of the degree to which the subjects have knowledge of or beliefs about their own knowledge and beliefs. In this essay, I have decided to idealize quite recklessly along the first dimension. I have, of course, allowed idealization along the second as well, but much less than the norm. The guiding intuition all along has been that, with respect to their attributions to themselves of knowledge and especially belief, the axiomatization should guarantee our subjects against certain kinds of epistemic/doxastic grief--*not that it should guarantee them all manner of epistemic/doxastic success.* Imagine a subject whose beliefs conform to our account. Such a subject will be under no pressure to change her beliefs about her beliefs--*no pressure, that is, stemming from* **conflicts** *between what she believes about what she believes and what she actually believes.* I assume, of course, that *falling short of "introspective omniscience"* **by itself** *generates no pressure, and no such* **conflicts**.

Let's return once again to **O** and **O' (= Y)**. **O** is a completely general schema to the effect, roughly, that our subject--Kimberly, say-- believes every true proposition. This is obviously bonkers. We allowed, however, that Kimberly does not believe the negation of **O**. This yielded **O'**, and **O'** simply denies that Kimberly believes things and also believes that she doesn't believe those things. It denies that Kimberly is subject to a certain kind of error of self-attribution--one might say *the basic kind* of such error. Note that by necessitation, Kimberly will of course believe that she is not thus subject to that kind of error. That is, she will believe, not that she has any real talent for doxastic self-attribution or introspection, but that she doesn't go around believing that she doesn't believe things she actually does believe.

Another way to see what's going on is to go farther than I have so far in intermixing belief and knowledge. At the moment the only two-operator schema I allow is to the innocuous effect that knowledge requires belief. In passing I mentioned Hintikka's argument against the epistemic version of **B**. **B** is sufficiently bizarre that one should not require even Scott to believe it; but what if we try out our trick on it? What about:

(-B-(-K-Kp --> p)) ?

What indeed? Let's transmogrify, using our recipe:

(-B(-K-Kp & -p))

It is not the case that Scott believes both not-p and that he does not know that he doesn't know that p. If he knew that p, he would not believe that not-p. (By **D** and the requirement that knowledge involves belief.) Of course, if he knew that he didn't know that p, he might very well believe that not-p. (Or not: he might be open-minded, have no opinion, with resppect to the question.) If he doesn't know that he doesn't know that p, he might still believe that not-p. After all, he just might not know that p, for instance, because he doesn't believe that p, but not know that he doesn't know it--for instance beeause he doesn't believe that he doesn't believe it. But if he were to believe that he doesn't know tbat he doesn't know that p--say, because he doesn't know that he doesn't believe that p--and yet still believe that not-p, then he would have reason for concern lest he be inconsistent, or of two minds about his attitude toward p (or its negation). And we have ruled out such worries.

Try another transform:

(-(B-K-Kp & B-p))

Either Scott doesn't believe not-p or he doesn't believe that "for all he knows", he knows that p--where, a la Hintikka, I'm reading '-K-q' as "for all Scott knows, q". So, imagine Scott believes that not-p. Then he had best not believe that for all he knows, he knows that p.

This trick works for the epistemic versions of **4** and **E** as well. No doubt looking at one of these will suffice. Let's do **4**, which--in its epistemic version, of eourse--was the most hotly contested of the "introspective axioms" originally proposed by Hintikka.

(-B(Kp & -KKp))

Scotty should not believe both he knows that p and that he doesn't know that he knows it. It is quite possible that Scotty know that p without knowing that he knows it. Remember, we reject **4**. But if he should believe that he knows that p, then it will not do to believe that he doesn't know that he knows it. Identifying Scott's being certain that p with his believing that he knows that p: if Scott is certain that p, then he doesn't believe that he doesn't know that he knows that p. (Although, again, he really might not know that he knows it.) He would not continue to be certain that p if he believed that he didn't know that he knew that p. Put otherwise: being certain that p requires not believing that for all you know you might not know that p.

8. SOME FINAL SCEPTICAL REMARKS.

Now to say a word about believing and knowing–in particular about believing. Believings and beliefs come in a wide variety of "modes". One talks of explicit and implicit beliefs, of conscious and unconscious beliefs, of occurrent ("active") and dispositional beliefs. These three dimensions/dichotomies are very likely independent, and there may be other such dimensions or dichotomies. To which of these, if any, is our 'B' operator supposed to correspond? Hintikka, for instance, clearly intends his 'B' and 'K' operators for what he calls "active belief" and "active knowledge." But he also seems to suppose that being active involves being conscious; that is, being an active belief involves being a belief of which the believer is conscious. Further, he argues–naturally enough–that it is only a certain mode (or modes) of believing for which various of his principles and rules are appropriate. Thus, for instance, the doxastic version of 4, that if one believes that p, one believes that one believes that p, holds of active, conscious beliefs. (He thus rules out of court–he thinks–references to Freud, self-deception, and the like.)

8.1. *On Belief States*

I can be no more than brief here, but it seems to me that a much more important "dichotomy" is that between two different conceptualizations of the role of belief. According to one conceptualization, the main locus or arena of beliefs is in thinking that is aimed at truth, that is, in "theoretical reasoning," considered in abstraction from the creature's possibilities of and requirements for action. This conceptualization leads quite naturally to foeussing on conscious beliefs, consciously arrived at, and thereby to focussing on language using creatures who can express their beliefs, including their beliefs about their own mental states. The other conceptualization might be called "functional"; Robert Stalnaker has called it "pragmatic-causal"[Stalnaker 85]. Here the main arena is aetion; the fundamental role of beliefs in the mental life of believers is as states that, together with desires and intentions, guide or direct or determine behavior. Roughly, to say that a subject believes that p is to say that if the subject were to desire that q, then he would be disposed to act in a way that would bring it about that q were it to be the case that p. This conceptualization of beliefs is essentially dispositional; within it, being active means playing the characteristic role of belief in an actual behavioral episode and has nothing whatsoever to do with being conscious–let alone with being linguistically expressible. Again, within this coneeptualization, neither language nor language users occupy any special privilged position of interest.

I take it that it is clear enough that a concern with "introspection" goes most naturally with the first of these two ways of thinking about belief. This is true even if one clearly distinguishes "introspective beliefs" from a subject's beliefs about its own mental states. Let me now clearly distinguish these two. The second has solely to do with the content of beliefs; a rough and ready characterization is simply this: the subject-matter of the creature's belief is about that creature's own mental states–including its own *present* mental states. Even here, and even in the case of beliefs about one's own

present mental-states, one can distinguish beliefs about one's own mental states in the "first-person mode" and in the "third-person" mode. So, for example, I might believe that the sixth oldest researcher in AI believes that p, without realizing that I *am* the sixth oldest researcher in AI. If so, I might be said to have a third-person belief about my own beliefs. In general, one can concoct examples in which a creature can have beliefs about itself without realizing that it is the very thing at which the beliefs in question are directed.

"Intropsective beliefs" on the other hand are beliefs about one's own mental states that are caused in a certain way (or ways), or which arise out of the functioning of one (or another) specific--though perhaps completely unspecified--cognitive mechanism called "introspection."[21] If one is thinking about beliefs from the "pragmatic-causal" perspective, it's hard to get excited about "introspective beliefs" unless one simply assumes that all beliefs that arise out of introspection are "in the first-person mode". But, then, what's crucial about them is that latter fact not their etiology.

Indeed, from within the "causal-pragmatic" or "functionalist" tradition, it's hard to get excited about epistemic/doxastic logic.[22]

8.2. *On the Contents of Beliefs*

Let me now say a word about the objects or contents of believings; that is, about beliefs. If the objects or contents of such mental states as believing and knowing are to be truth-valuable--as they are represented as being in all standard epistemic and doxastic logics, then they had best make or correspond to or just *be* determinate claims upon reality. Sentences--sentences types--of natural languages precisely **do not** correspond to or make such claims. Sentences--better, well formed formulae--of standard logical languages, by tacit conventions of interpretation or of intended range of applicability, are supposed to make such determinate claims. That is, such sentences are supposed to be eternal: any statement-making utterance of such a sentence yields the same propositional upshot, makes the same determinate claim upon the world. (Of course, the worlds in question are typically conceived of as mathematical structures: that is, as consisting of eternal or timeless objects standing in certain timeless relations among themselves.) So if we imagine a subject whose beliefs are mediated (carried) by, as well as being expressible in, sentences of such a formal language, that is, if we imagine the subject's believing that p as

---

[21]Of course, given this characterization, it is really an open question whether there are any introspective beliefs. I think that there are; but that the members of only very few species can have them. I used to think that only the members of language-using species could; I am now prepared to be more liberal and include those species which manifest a certain kind and degree of social or group organization. Unfortunately, I can't characterize this kind or that degree; nor do I have any good argument as to the necessity of the alluded to condition. But then again no one has ever told me of what introspection really consists.

[22]**NOTA BENE**: from within the "pragmatic-causal" world picture, what's crucial seems to be the "first-person" mode; for it is self-attribution in that mode which guides action. Or, perhaps one should say that what is crucial is the relation between the first-person and the third-person modes of self-attribution See, e.g., [Perry 85].

involving the subject's saying--to himself, for example--a sentence of such a language, then we *might* have no trouble convincing ourselves that the content of such a subject's beliefs are transparent, completely accessible, to that subject. This is again precisely what we cannot imagine even if we follow this language-involving conception of belief but think instead of our subject thinking to itself in (using) sentences of some natural language. (In all of this, I am assuming complete semantic competence--though, of course, without having a complete theory of what constitutes such competence. At any rate, I am assuming--for the sake of argument--that such competence consists, at least, in the subject's knowing what any sentence of his language "means"; so, in the case of a language with only eternal sentences, in knowing for every sentence, what claim is made by that sentence--what the world would have to be like for that sentence to be true.) If we deny that believing is essentially language involving, it is harder still to see why or even how the content of a subject's beliefs should be transparent to that subject.

Moreover, if we are working within the functionalist paradigm, we will see that--in so far as we are interested in generalizations across subjects or across time and changing circumstances--our primary interest will be in a notion of content under which contents are not truth-valuable and do not correspond to determinate claims upon reality. Note, well, that I speak of "content", not of "object; I don't think there is a useful sense in which the meanings of non-eternal sentences are objects of belief. We shall, that is, be interested in a notion of content such that (e.g.) when both Scott and Kimberly say to themselves, "There's no milk in the fridge," even if at different times and locations, and the like, the mental states that such imagined sayings indicate have the same content. For if both desire to drink some milk, or even if both desire that there be some milk in the fridge, then they would be disposed to act in such a way as to bring it about that there would be milk in the fridge were it the case that there was as yet no milk in the fridge. Just as the truth-valuable contents of their mental states are different, so too are the contents (objects) of their desires, both their desires *to* (drink some milk) and their desires *that* (there be milk in one's fridge). Note, too, the talk of "act in such a way." The way or ways in question can only be characterized at a level of abstraction or generality that cuts across the differences in the actual circumstances of Scott and Kimberly and cuts across them in a way correlative to that in which the sameness of their mental states cuts across the differences in the truth-valuable contents/objects of their beliefs. Much mischief has been wrought by failure to distinguish these two different conceptions of content [Barwise and Perry 83]. Finally and to repeat: from within the functionalist perspective, it is the second notion of content that is crucial, or--again--the relation between the two notions. Hence again, what interest could there be, from within such a conceptualization, in standard epistemic/doxastic logics--formalisms which, at least standardly, take it that the proper objects of belief, within the logic, are truth-valuable and propositional?[23]

[23]Here I should remind the reader that within epistemic and doxastic logics, belief and knowledge don't really get treated as relations to propositions; that is, such logics are to be contrasted with theories--say, first or higher order theories--of the relations in question. In the context of these intensional logics, the relations are metatheoretic epiphenomena, arising out of a particular heuristic for understanding a particular model-theoretic treatment.

It may be, then, that to take epistemic/doxastic logics seriously, one must both be working from within that conceptualization of cognitive states according to which they are either essentially or importantly language involving and, further, conceive of the language(s) in question on the model of standard formal languages, as consisting, that is, of eternal sentences only. This *could* be taken as an argument to the effect that the proper home of epistemic/doxastic logic is theoretical computer science—precisely the locus of its greatest current vitality.

References

[1] Anderson, A. and Belnap, N.
Entailment, Volume I
Princeton University Press, Princeton, NJ, 1975

[2] Barwise, J. and Perry, J.
Situations and Attitudes
Bradford Books, MIT Press, Cambridge, MA, 1983

[3] Binkley, R.
The Surprise Examination in Modal Logic
Journal of Philosophy 65 (1968) pp. 127-136

[4] Fagin, R. and Halpern, J.
Belief, Awareness, and Limited Reasoning
In **Proceedings of the Ninth International Joint Conference on Artificial Intelligence**
1985

[5] Fagin, R. and Vardi, M.
An Internal Semantics for Modal Logic
In **Proceedings of the Seventeenth ACM Conference on Theory of Computing**
1985

[6] Halpern J. and Moses, Y.
Knowledge and Common Knowledge in a Distributed Environment
In **Proceedings of the Third ACM Conference on Principles of Distributed Computing**
1984

[7] Harrison, C.
The Unanticipated Examination in view of Kripke's Semantics for Modal Logic
In **Philosophical Logic**, ed. J, Davis, D. Hockney, and W. 8}Wilson, D. reidel, Dordrecht, 1969

[8] Hintikka, J.
Knowledge and Belief: An Introduction to the Logic of the Two Notions
Cornell University Press, Ithaca, NY, 1962

[9] Hintikka, J.
Impossible Possible Worlds Vindicated
J. of Philosophical Logic 4 (1975), pp. 475-484

[10] Hintikka, J.
"Knowing that one knows" Reviewed
Synthese 21 (1970)

[11] Lemmon, E. J., in collaboration with D. Scott
The "Lemmon Notes": An Introduction to Modal Logic

ed. K. Segerberg
Blackwell, Oxford, 1977

[12] Lewis, C. I.
A Survey of Symbolic Logic
University of California Press, Berkeley, CA, 1918
(reprinted, with excisions, by Dover, 1961)

[13] Perry, J.
Perception, Action, and the Structure of Believing
to appear in a **Festschrift for Paul Grice**, ed. R. Grandy
and R. Warner, Oxford University Press, Oxford, 1985

[14] Prior, A. N.
A Budget of Paradoxes
in **Objects of Thought**, ed. P. T. Geach and A. J. P. Kenny,
Oxford University Press, Oxford, 1971

[15] Stalnaker, R.
Inquiry
Bradford Books, MIT Press, Cambridge, MA, 1984